

Information Causality is a Special Point in the Dual of the Gray-Wyner Region

Salman Beigi

School of Mathematics, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran

Amin Gohari

*Department of Electrical Engineering, Sharif University of Technology, Tehran, Iran
School of Computer Science, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran*

Information Causality contributes to the program of deriving fundamentals of quantum theory from information theoretic principles. It puts restrictions on the amount of information learned by a party (Bob) from the other party (Alice) in a one-way communication scenario as follows. Bob receives an index b , and after a one-way communication from Alice, tries to recover a part of Alice's input. Because of the possibility of cloning, this game in its completely classical form is equivalent to one in which there are several Bobs indexed by b , who are interested in recovering different parts of Alice's input string, after receiving a *public* message from her. Adding a *private* message from Alice to each Bob, and assuming that the game is played many times, we obtain the Gray-Wyner problem for which a complete characterization of the achievable region is known. In this paper, we first argue that in the classical case Information Causality is only a single point in the dual of the Gray-Wyner region. Next, we show that despite the fact that cloning is impossible in a general physical theory, the result from classical world carries over to any physical theory provided that it satisfies a new property. This new property of the physical theory is called 'Accessibility of Mutual Information' and holds in the quantum theory. We conclude that the Gray-Wyner region completely characterizes all the inequalities corresponding to the game of Information Causality. In other words, we provide infinitely many inequalities that Information Causality is only one of them.

In the second part of the paper we show that Information Causality leads to a non-trivial lower bound on the communication cost of simulating a given non-local box when the parties are allowed to share *entanglement*. We also consider the same problem when the parties are provided with preshared randomness.

I. INTRODUCTION

Non-locality is arguably the most fundamental feature of quantum physics. Bell's theorem [1], as verified by experiments [2], states that there are correlations in nature that cannot be explained by local realistic (classical) theories. Bell's inequalities restrict the strength of classical correlations, while in the quantum theory correlations are characterized by Tsirelson's bounds [3]. The latter bounds, however, heavily rely on the seemingly ad hoc postulates of quantum mechanics. On the other hand, non-locality, the property that makes physical theories to depart from the classical ones, is a fundamental feature of nature rather than quantum mechanics by itself. Tsirelson's bounds then do not provide a satisfactory answer to the problem of quantifying non-locality.

Recently, there has been a stream of works to understand non-locality from more fundamental principals. No-signaling as the first such principal does not describe correlations of quantum physics since non-signaling PR-boxes [4] maximally violate the Tsirelson bound (and then Bell's inequality) for the CHSH expression [5] and do not seem to be physical. Nevertheless, the recently proposed principal of Information Causality [6], a generalization of no-signaling, exactly gives Tsirelson's quantum bound for the CHSH. Thus this is a natural question whether Information Causality or other information theoretic principals can further our understanding of non-locality.

A. Information Causality

Let us briefly explain the game of Information Causality. Alice receives the bit-string $\vec{a} = (a_1, \dots, a_N)$ consisting of i.i.d. random bits, and Bob gets an index $1 \leq b \leq N$. Bob's goal is to output a_b upon receiving a classical message x from Alice. Assuming that β_i is Bob's guess of a_i when $b = i$, Information Causality states that

$$H(x) \geq \sum_{i=1}^N I(a_i; \beta_i | b = i). \quad (1)$$

It turns out one can rewrite this inequality in terms of entropies as follows:

$$H(x) + \sum_{i=1}^N H(a_i|\beta_i, b=i) \geq H(\vec{a}). \quad (2)$$

Despite its success in characterizing certain regions of non-local correlations, Information Causality seems specifically designed for the CHSH. The underlying game of Information Causality is not the general one-way communication scenario one could consider. The parallel repetition of the game cannot be expressed as a special instance of the game itself, as one would expect from an information theoretic concept (information theoretic concepts are mainly defined in an asymptotic sense). Moreover, the final inequality of Information Causality seems arbitrary and one may ask about other combinations of the terms appearing in its expression. Here we argue that the individual terms are indeed the right ones, but the combination in Information Causality is only a special one.

Consider the game of Information Causality in its completely classical form. Classicality enables us to assume that there are N Bobs instead of one. We denote these N Bobs by $\text{Bob}_1, \dots, \text{Bob}_N$. The goal of Bob_i is to find a_i . Moreover, we may assume that shared randomness is indeed shared amongst Alice and all Bobs, and all of them receive the message x . Then the first term $H(x)$ of (2) is the amount of information that is sent to all Bobs; the second term $H(a_i|\beta_i, b=i)$ expresses the remaining uncertainty of Bob_i about a_i . We can interpret this as the average number of extra bits that Alice needs to privately send to Bob_i to enable the recovery of a_i by this party if they were to play multiple copies of this game in parallel (the Slepian-Wolf theorem). Since $\text{Bob}_1, \dots, \text{Bob}_N$ altogether can recover the string \vec{a} , the total flow of information from Alice should be at least $H(\vec{a})$ by the cut-set bound. That is, the sum of the terms on the left hand side of (2) should dominate $H(\vec{a})$. This gives a new proof of Information Causality in the classical world.

The above game among Alice and the multiple copies of Bob has a similar setup to the Gray-Wyner problem [7]. The Gray-Wyner problem will be rigorously explained later, but roughly speaking it is defined as follows. Alice sends a public message x to all Bobs and afterwards a private message to each Bob_i . The goal of Bob_i is to recover a_i with a vanishing probability of error (see Fig. 1). Let R_0 denote the information rate of the public message, and (R_1, \dots, R_N) denote the rate of the private messages to $\text{Bob}_1, \dots, \text{Bob}_N$. The Gray-Wyner region explicitly characterizes the set of tuples (R_0, R_1, \dots, R_N) for which it is possible to satisfy the demands of $\text{Bob}_1, \dots, \text{Bob}_N$. This implies that the rates $R_0 = H(x)$ and $R_i = H(a_i|\beta_i, b=i)$ have to lie in the Gray-Wyner region when the Information Causality game is played in the classical world.

A main contribution of our work is that despite the fact that the cloning of Bob is impossible in a general physical theory, the tuple $(H(x), H(a_1|\beta_1, b=1), \dots, H(a_N|\beta_N, b=N))$ would still fall in the Gray-Wyner region if the physical theory satisfies a new property (besides the ones in [6]). This new property of the physical theory is called the Accessibility of Mutual Information and holds in the quantum theory. For any physical theory satisfying these properties, there are infinitely many inequalities originated from the characterization of the Gray-Wyner region; Information Causality is only one of them. In fact, the Gray-Wyner region completely characterizes all the inequalities corresponding to the game of Information Causality in the following sense. On one hand, the tuple $(H(x), H(a_1|\beta_1, b=1), \dots, H(a_N|\beta_N, b=N))$ has to be in the Gray-Wyner region. On the other hand, any point in the Gray-Wyner region is achievable, meaning that it can be obtained through a communication scheme in the classical world.

B. Simulation of non-local correlations

Quantifying the amount of classical communication required for simulating non-local correlations is the other well-studied approach, besides Bell's theorem, in the theory of non-locality (see e.g. [8–15]). A well-known result in this direction says that any bipartite correlation coming from one bit of entanglement (an EPR pair) can be realized in the classical world by only one bit of communication [8].

Here we introduce a novel application of Information Causality in the problem of simulating non-local correlations using classical one-way communication. We show that Information Causality leads to a non-trivial lower bound on the communication cost of simulating a given non-local box when the parties are allowed to share entanglement.

We also do have a non-technical contribution if one is interested in the communication cost of simulating a given non-local box when the parties only share common *randomness*. We comment that information theorists who have been interested in the area of control have independently studied the same problem in a different context. To the best knowledge of the authors, however, all previous results in quantum information attack the problem from a communication complexity point of view, and not information theoretic. The communication complexity formulation of the problem turns out to be a very difficult one. However, information theoretic formulation of the problem looks at the limits of the problem and takes the advantage of laws of large numbers. Connecting these two lines of research, we report a formula that gives an *exact* expression for the optimal amount of communication needed for non-local

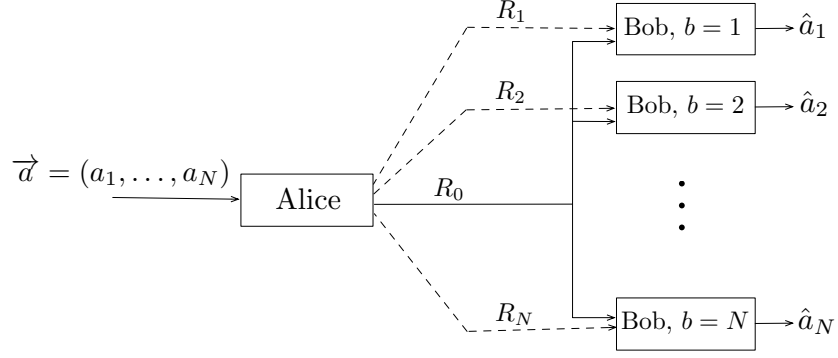


FIG. 1: The Gray-Wyner game consists of $N + 1$ players, Alice and N Bobs who are indexed by $b = 1, \dots, N$. Alice receives the i.i.d. copies of (a_1, \dots, a_N) , sends public information at rate R_0 to all Bobs and private information at rate R_i to Bob $_i$. The goal of Bob $_i$ is to recover a_i .

simulation given preshared randomness. It should be also noted that the information theoretic characterization of the communication cost serves as a lower bound on the communication complexity characterization of the bound, because the former setup considers asymptotic behaviors and is more relaxed.

II. REVIEW OF INFORMATION CAUSALITY

In this paper we mainly adopt the notation used in [6]. Alice receives the bit-string $\vec{a} = (a_1, \dots, a_N)$ consisting of i.i.d. random bits, and Bob gets an index $1 \leq b \leq N$. Their goal is that Bob after receiving a classical message x from Alice, outputs a_b . Assuming that β_i is Bob's guess of a_i when $b = i$, Information Causality states that

$$H(x) \geq \sum_{i=1}^N I(a_i; \beta_i | b = i). \quad (3)$$

Before getting into our main results and discussion of the Information Causality in the next section, we begin by slightly generalizing its game. The game of Information Causality, as formulated in [6], is a special one-way communication problem. One can generalize it by assuming that instead of outputting a single bit of Alice, Bob may want to compute some function of Alice's input a , and his input b : $f(a, b)$. To fit this new scenario into the previous setup, assume that b takes values $1, 2, \dots, N$, and note that Alice may replace her input a with the string $(a_1, a_2, \dots, a_N) = (f(a, 1), f(a, 2), \dots, f(a, N))$ which by slightly abuse of notation we represent by \vec{a} . Note that \vec{a} is a sufficient statistic from the perspective of Alice, and she can use it instead of a . Bob's goal then is to find a_b , the b -th coordinate of \vec{a} as before. The difference, however, is that a_i 's are no longer i.i.d. and can be correlated. As a summary, the problem of Bob aiming to compute a function of Alice's input can be converted to the problem of a_i 's being correlated. Observe that by considering correlated inputs a_1, \dots, a_n the parallel repetition of the game is itself.

Now we should seek for an adjustment of the proof of (3) in [6] that admits correlated a_i . Following the proof, we find that the independence of a_i 's is used in [6] where the term $I(a_{j+1}, a_{j+2}, \dots, a_N; a_j)$ is dropped. When a_i 's are correlated, we have to put these penalty terms back. A careful bookkeeping of the penalty terms then gives us $\sum_{i=1}^{N-1} I(a_{i+1}, a_{i+2}, \dots, a_N; a_i)$ which is equal to $\sum_{i=1}^N H(a_i) - H(\vec{a})$. Adding this to equation (3) we obtain

$$H(x) \geq \sum_{i=1}^N I(a_i; \beta_i | b = i) - \sum_{i=1}^N H(a_i) + H(\vec{a}).$$

If we expand the mutual information term $I(a_i; \beta_i | b = i)$ as $H(a_i) - H(a_i | \beta_i, b = i)$, we get a simpler representation of the above equation

$$H(x) + \sum_{i=1}^N H(a_i | \beta_i, b = i) \geq H(\vec{a}). \quad (4)$$

In the rest of the paper we still call the above inequality for correlated a_i 's, Information Causality.

Before finishing this section let us review the conditions given in [6] under which (4) holds in a given physical theory. Information Causality holds *if* a symmetric and non-negative mutual information can be defined for every two systems in the theory that obeys the following three properties directly quoted from [6]:

(1) *Consistency*: If the subsystems A and B are both classical, then $I(A; B)$ should coincide with Shannon's mutual information.

(2) *Data Processing Inequality*: Acting on one of the parts locally by any state transformation allowed in the theory cannot increase the mutual information. I.e., if $B \rightarrow B'$ is a permissible map between systems, then $I(A; B) \geq I(A; B')$.

(3) *Chain Rule*: There exists a *conditional mutual information* $I(A; B|C)$ such that the following identity is satisfied for all states and triples of parts: $I(A; B, C) = I(A; C) + I(A; B|C)$.

Since in this paper we are writing Information Causality in terms of entropy (equation (4)) rather than mutual information (equation (3)) we need another property defining the entropy.

(4) *Mutual Information and Entropy*: For every system A there exists a non-negative number $H(A)$ that equals the Shannon entropy of A if it is classic. Furthermore, for subsystems A and B we have $I(A; B) = H(A) + H(B) - H(A, B)$. And by $H(A|B)$ we mean $H(A, B) - H(B)$.

Remark II.1. In [16] and [17] equation (4) has been derived from some postulates directly on the entropy rather than mutual information. The above four properties thus can be replaced with two.

We will impose yet another property on mutual information called the ‘Accessibility of Mutual Information’ (AMI).

(5) *Accessibility of Mutual Information (AMI)*: Consider arbitrary subsystems A and B where A is classical. Let $(A_1, B_1), (A_2, B_2), \dots, (A_n, B_n)$ denote n independent copies of (A, B) . Then for any $\epsilon > 0$, there exists some n and a local state transformation $(B_1, \dots, B_n) \rightarrow e_n$, such that e_n is classical and

$$\frac{1}{n} I(A_1 \dots A_n; e_n) \geq I(A; B) - \epsilon.$$

Properties (1), (3) and (4) have nothing to do with the space of valid state transformations in the underlying physical theory. Property (2), on the other hand, restricts this space, and can always be satisfied by putting enough constraints on the set of valid maps between systems. For instance data processing inequality becomes trivial in a physical theory whose only valid state transformation is the identity map. Thus to avoid such obscure examples, an information theoretic approach to study physical theories has to provide another postulate, besides (1)-(4), about the richness of the space of valid state transformations. From this point of view AMI which ensures the existence of certain maps, is a natural property. Moreover, this is the property that formulates our intuition of mutual information, and otherwise the function of mutual information on non-classical systems has no tangible meaning.

Observe that AMI holds in quantum physics because the Holevo outer bound on the accessible information is asymptotically achievable by the pretty good measurement.

III. INFORMATION CAUSALITY AND THE GRAY-WYNER PROBLEM

As discussed in Sec. IA the game of Information Causality is closely related to the Gray-Wyner problem. In the latter, there are one encoder (Alice) and N decoders ($\text{Bob}_1, \dots, \text{Bob}_N$). Alice is observing i.i.d. copies of $\vec{a} = (a_1, \dots, a_N)$, where $a_i \in \mathcal{A}_i$ takes discrete values, and can send a public message at rate R_0 to all Bobs, and N private messages at rates R_1, R_2, \dots, R_N (at rate R_i to Bob_i). The goal is for Bob_i to recover the i.i.d. copies of a_i with probability of error converging to zero as the number of i.i.d. observations goes to infinity (see Fig. 1). The Gray-Wyner region \mathcal{R} is defined to be the set of *achievable* rate vectors (R_0, R_1, \dots, R_N) , i.e., $(R_0, R_1, \dots, R_N) \in \mathcal{R}$ if by sending public and private information at rates R_0 and R_1, \dots, R_N respectively, Bobs' demands can be fulfilled.

The set \mathcal{R} of achievable rate vectors for the Gray-Wyner problem is completely characterized. $(R_0, R_1, \dots, R_N) \in \mathcal{R}$ if and only if there exists an auxiliary random variable e such that

$$R_0 \geq I(\vec{a}; e), \quad (5)$$

$$R_i \geq H(a_i|e), \quad 1 \leq i \leq N. \quad (6)$$

Note that e is classic and is determined by the conditional probability vector $p(e|\vec{a})$. Moreover, without loss of generality we may assume that e takes values in a discrete set of size $(\prod_i |\mathcal{A}_i| + 2)$ [18]. This enables us to explicitly compute \mathcal{R} .

Let us now consider the Information Causality game in the classical case and consider the $(N+1)$ -tuple

$$(H(x), H(a_1|\beta_1, b=1), H(a_2|\beta_2, b=2), \dots, H(a_N|\beta_N, b=N)). \quad (7)$$

We claim that this vector is in the set \mathcal{R} . One can see this by noting that the Gray-Wyner problem is indeed the Information Causality game allowed to be played many times. Equivalently, for the choice of $e = (x, c)$ where c is the common randomness shared between the two parties one has

$$H(x) \geq I(\vec{a}; e), \quad (8)$$

$$H(a_i|\beta_i, b=i) \geq H(a_i|e), \quad 1 \leq i \leq N. \quad (9)$$

The first equation comes from the fact that c is independent of \vec{a} , and the second one is the data processing inequality. Now adding up these equations, we obtain

$$\begin{aligned} H(x) + \sum_{i=1}^N H(a_i|\beta_i, b=i) &\geq I(\vec{a}; e) + \sum_{i=1}^N H(a_i|e) \\ &= H(\vec{a}) - H(\vec{a}|e) + \sum_{i=1}^N H(a_i|e) \\ &\geq H(\vec{a}), \end{aligned}$$

where in the last line we use the subadditivity of entropy. Thus we obtain the Information Casualty bound in the classic world. In general, summation of (8) and (9) with weights $w_i \geq 0$ gives

$$w_0 H(x) + \sum_{i=1}^N w_i H(a_i|\beta_i, b=i) \geq w_0 I(\vec{a}; e) + \sum_{i=1}^N w_i H(a_i|e). \quad (10)$$

Now if we go beyond the classic world and for instance consider the quantum theory, it is no longer clear that the $(N+1)$ -tuple of equation (7) would always fall in the set \mathcal{R} . If we attempt to mimic the above proof, shared randomness should be replaced with the shared entanglement. If we let B to be Bob's subsystem of the shared quantum state, then we need to identify the auxiliary random variable e by $e = (x, B)$. In this case, inequalities (8) and (9) do hold because firstly, B is independent of \vec{a} and secondly, the data processing inequality is still available. However, our choice of e is *not* classic anymore while the Gray-Wyner region is defined using classical auxiliary random variable e . Furthermore, because cloning is impossible in the quantum theory, we can no longer consider several Bobs and the conceptual link with the Gray-Wyner problem is lost.

One of the main contributions of this paper is that the vector of equation (7) does indeed fall in \mathcal{R} in the quantum world. More generally, we show that the $(N+1)$ -tuple falls in \mathcal{R} in any physical theory that satisfies properties (1-5) mentioned in Sec. II. This is made rigorous and proved in Section IV. But now we discuss some implications of this result.

A. Examples and implications

In an arbitrary physical theory if the $(N+1)$ -tuple (7) falls in \mathcal{R} , equation (10) holds for some $p(e|\vec{a})$. In other words, for arbitrary coefficients $w_i \geq 0$ we have

$$w_0 H(x) + \sum_{i=1}^N w_i H(a_i|\beta_i, b=i) \geq w_0 H(\vec{a}) + \inf_{p(e|\vec{a})} \left(-w_0 H(\vec{a}|e) + \sum_{i=1}^N w_i H(a_i|e) \right). \quad (11)$$

The above equation can be thought of as a generalization of Information Causality (4). This generalization is strictly stronger than the original one. For instance, consider the following scenario where Bob receives either an index $1 \leq b \leq N$ or two indices $1 \leq b_1, b_2 \leq N$. In the former case he wants to recover a_b , and in the latter both a_{b_1} and a_{b_2} . We claim that (4) is loose for this game while we offer much stronger inequalities. To see this note that (4) can be written as

$$H(x) + \sum_{i=1}^N H(a_i | \beta_i, b = i) + \sum_{i,j=1}^N H(a_i, a_j | \beta_i, \beta_j, (b_1, b_2) = (i, j)) \geq H(\vec{a}).$$

However, in our generalization we can set the coefficient $w_0 = 1$, and $w_i = 1$ for $i \geq 1$ when the i -th term corresponds to the case of Bob receiving a single index, and $w_i = 0$ otherwise. In this way we get away with the terms of the form $H(a_i, a_j | \beta_i, \beta_j, (b_1, b_2) = (i, j))$ because their coefficient is zero. We obtain the strictly stronger inequality

$$\begin{aligned} H(x) + \sum_{i=1}^N H(a_i | \beta_i, b = i) &\geq \inf_{p(e|\vec{a})} I(\vec{a}; e) + \sum_{i=1}^N H(a_i | e) \\ &\geq H(\vec{a}), \end{aligned}$$

where in the last step we have used the subadditivity of entropy.

To illustrate the benefit of expressing the game of Information Casualty in terms of the Gray-Wyner region we consider another example. Assume that $N = 2$ and Bob receives an index $b \in \{1, 2\}$ and wants to recover a_b . We assume that random variables a_1, a_2 are correlated. By the original Information Causality equation (4), we have

$$H(x) + H(a_1 | \beta_1, b = 1) + H(a_2 | \beta_2, b = 2) \geq H(a_1, a_2).$$

Now we claim that if $H(x)$ is sufficiently small, the above inequality is loose. This is clear when $H(x) = 0$ because a_1, a_2 are correlated. In general the inequality is loose whenever $H(x)$ is less than $J(a_1, a_2)$ the Wyner's common information between a_1 and a_2 [28]. To prove the above claim we exploit a known result about the two receiver Gray-Wyner problem (see pp. 367-368 of [19]) saying that the minimum value of R_0 when $R_0 + R_1 + R_2 = H(a_1, a_2)$ is the Wyner's common information. The claim follows from this result and the fact that the triple $(R_0 = H(x), R_1 = H(a_1 | \beta_1, b = 1), R_2 = H(a_2 | \beta_2, b = 2))$ is in the Gray-Wyner region. Thus, when $R_0 = H(x) < J(a_1, a_2)$, we should look into the full characterization of the Gray-Wyner region, which can be expressed in terms of all inequalities of the form (11).

When Alice's inputs a_1, a_2, \dots, a_N are mutually independent, the Gray-Wyner region is fully characterized by the single inequality with all weights $w_i = 1$, thus our main contribution is in the case where a_i 's are not mutually independent.

IV. PROOF OF THE MAIN RESULT

In this section we prove the following theorem.

Theorem IV.1. *Fix a strategy for the Information Causality game in a physical theory that satisfies properties (1-5) of Sec. II including Accessibility of Mutual Information. Then the $(N + 1)$ -tuple*

$$(H(x), H(a_1 | \beta_1, b = 1), H(a_2 | \beta_2, b = 2), \dots, H(a_N | \beta_N, b = N))$$

falls in the Gray-Wyner region \mathcal{R} corresponding to $p(a_1, a_2, \dots, a_N)$ given by equations (5) and (6).

Remark IV.2. *As in [6], we would like to emphasize that all of the quantities showing up in the statement of the theorem “do not involve the details of a particular physical model but are fully determined by Alice's and Bob's input bits and Bob's output.”*

Proof. Let $u = (x, B)$. We claim that

$$H(x) \geq I(\vec{a}; u), \tag{12}$$

$$H(a_i | \beta_i, b = i) \geq H(a_i | u), \quad 1 \leq i \leq N. \tag{13}$$

The first inequality was proved in [6] and holds because B is independent of \vec{a} . The second one comes from the data processing inequality and the fact that b is independent of (\vec{a}, x, B) .

For a physical system α (random variable in the classical case) we denote α^n to be n independent copies of α . Fix an arbitrary $\epsilon > 0$. The rest of the proof can be divided into two parts.

(I) There exists a natural number n and a random variable e_n determined by the conditional distribution $p(e_n|\vec{a}^n)$ such that

$$I(\vec{a}; u) \geq \frac{1}{n} I(\vec{a}^n; e_n), \quad (14)$$

$$H(a_i|u) \geq \frac{1}{n} H(a_i^n|e_n) - \epsilon, \quad 1 \leq i \leq N. \quad (15)$$

(II) There exists a random variable e^* (again determined by $p(e^*|\vec{a})$) such that

$$\begin{aligned} \frac{1}{n} I(\vec{a}^n; e_n) &= I(\vec{a}; e^*), \\ \frac{1}{n} H(a_i^n|e_n) &\geq H(a_i|e^*), \quad 1 \leq i \leq N. \end{aligned}$$

(I) and (II) together with (12) and (13) imply that

$$\begin{aligned} H(x) &\geq I(\vec{a}; e^*), \\ H(a_i|\beta_i, b=i) &\geq H(a_i|e^*) - \epsilon, \quad 1 \leq i \leq N. \end{aligned}$$

In other words, $(H(x), H(a_1|\beta_1, b=1) + \epsilon, H(a_2|\beta_2, b=2) + \epsilon, \dots, H(a_N|\beta_N, b=N) + \epsilon)$ belongs to \mathcal{R} . Since $\epsilon > 0$ is arbitrary and the set \mathcal{R} is closed [29], we obtain the desired result.

We prove (I) here and leave the proof of (II), which follows from standard tricks in information theory, for Appendix A.

AMI for the pair (\vec{a}, u) implies that there exist a natural number n and a local processing $u^n \rightarrow e_n$ where e_n is classical such that

$$\frac{1}{n} I(\vec{a}^n; e_n) \geq I(\vec{a}; u) - \epsilon. \quad (16)$$

Equation (14) then follows from the data processing inequality: $I(e_n; \vec{a}^n) \leq I(u^n; \vec{a}^n) = nI(u; \vec{a})$.

Proof of (15) is easier if we rewrite it as

$$\frac{1}{n} I(a_i^n; e_n) \geq I(a_i; u) - \epsilon.$$

By the data processing inequality [30] we have $I(\vec{a}^n; u^n|a_i^n) \geq I(\vec{a}^n; e_n|a_i^n)$. Then using the fact that \vec{a}^n contains a_i^n we obtain

$$I(\vec{a}; u) - I(a_i; u) \geq \frac{1}{n} (I(\vec{a}^n; e_n) - I(a_i^n; e_n)),$$

or equivalently

$$I(\vec{a}; u) - \frac{1}{n} I(\vec{a}^n; e_n) \geq I(a_i; u) - \frac{1}{n} I(a_i^n; e_n).$$

But by (16), the left hand side is at most ϵ , and we are done.

V. SIMULATION OF NON-LOCAL CORRELATIONS

Finding the amount of classical communication required to simulate non-local correlations is a well-known method to quantify non-locality (see e.g. [8–15]). In this section we argue that Information Causality could enhance our understanding of this problem. Furthermore, because the problem of simulating non-locality is important on its own, we also aim to encourage study of quantification of non-locality from an information theoretic perspective, rather than a communication complexity one.

Given a certain non-local box, we consider the problem of the minimum amount of communication from Alice to Bob required to simulate the box. It is possible to consider multi-round communication scenarios as well, but unless stated otherwise we consider only one-way communication schemes. Here we assume that Alice and Bob have infinite shared randomness, and similarly the entanglement-assisted version of this problem can be considered. We prove that Information Causality leads to a lower bound on the communication cost of simulating a given non-local box when the two parties are provided with preshared entanglement.

Take an arbitrary non-local box, and consider the Information Causality game when Alice and Bob are provided with copies of this box as a resource. (This is indeed the scenario considered in [6].) Moreover, take a certain scheme where a message x is transmitted and a random variable β_i is constructed by Bob. Assume that the two parties use k copies of the box in this protocol. We would like to simulate this scheme by two new parties, say Alice' and Bob', who have only access to quantum entanglement as their resources at the outset.

Let C_{box} be the entanglement-assisted communication cost of simulating the non-local box. Alice' and Bob' can simulate the scheme of Alice and Bob by first sending kC_{box} bits from Alice' to Bob' to simulate the k boxes, and then $H(x)$ bits to simulate the message that was passed from Alice to Bob. This enables Bob' to faithfully simulate β_i . Now it is legitimate to write the Information Causality principle for the simulated protocol because it is happening in the quantum world. The total size of transmitted message is $kC_{\text{box}} + H(x)$. Therefore,

$$kC_{\text{box}} + H(x) \geq \sum_{i=1}^N I(a_i; \beta_i | b = i).$$

As an example let us consider the imperfect PR-box with bias ϵ . That is for binary inputs a, b and outputs x, y , $x \oplus y = ab$ with probability $\frac{1+\epsilon}{2}$. The scheme provided in [6] for $N = 2^n$ uses $k = 2^n - 1$ of these boxes, and the right hand side of the above equation is computed to be $2^n (1 - h(\frac{1+\epsilon^n}{2}))$. This gives us the following equation which results in a lower bound on C_{box} .

$$(2^n - 1)C_{\text{box}} + 1 \geq 2^n (1 - h(\frac{1+\epsilon^n}{2})).$$

Computing this lower bound for all n and taking the optimal one for every ϵ , we obtain the plot of Fig. 2. We see that the lower bound is equal to one at $\epsilon = 1$, thus it has to be tight at this point. By [6], the above lower bound (for n converging to infinity) would also be tight at the other end point $\epsilon \leq \frac{1}{\sqrt{2}} \simeq 0.707$. However, it may be loose in between because firstly we have considered the specific scheme of [6] for using boxes, and secondly this lower bound holds more generally for any physical theory satisfying properties of mutual information given in [6] and not only for quantum physics. Nonetheless, we would like to highlight that the lower bound at $\epsilon = 1$ is tight in any such physical theory, as shown in the figure.

Although this method for finding a lower bound on the entanglement-assisted communication cost of simulating non-local boxes works for any box, we only have the specific example of imperfect PR-boxes. This is because to the best knowledge of authors PR-boxes (and their generalizations [20]) are the only example of non-local boxes for which a relatively efficient scheme for solving communication problems is known.

A. Non-local box simulation from an information theoretic perspective

In the previous section we considered the problem of simulation of non-local boxes when the two parties share entanglement. In this section, we are interested in the same problem when the parties share classical common randomness. Our purpose here is to draw connections between the problem at hand, and a control problem studied by information theorists. We will report a formula that gives an *exact* expression for the optimal amount of communication needed for non-local simulation given preshared randomness. It should be also noted that the information theoretic characterization of the communication cost serves as a lower bound on the communication complexity characterization of the bound, because the former setup considers asymptotic behaviors and is more relaxed.

Information theorists have looked at the problem of simulating non-local correlations in a different context without linking it to quantum physics. Indeed this problem is related to the problem of coordinating distributed controllers to carry out some joint action (see [21, 22]). Now, if one were to formulate the problem of simulating non-local correlations in an information theoretic framework, it would go along the following lines: we are interested in simulating many independent copies of a non-local box, and ask for the minimum communication needed *per* box. Interestingly this formulation coincides with the formulation of the control problem.

More precisely, take an arbitrary bipartite box $p(x, y | a, b)$. Random variables x and y are not meaningful unless we specify a joint distribution on the pair (a, b) . So let us fix a distribution $p(a, b)$ on a, b as well. Now consider the

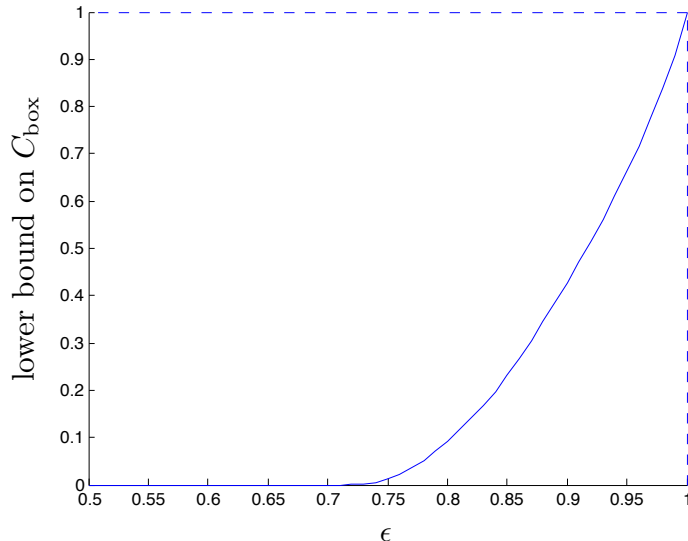


FIG. 2: A lower bound on the entanglement-assisted one-way communication cost of simulating imperfect PR-boxes with parameter ϵ , where $p(x \oplus y = ab) = \frac{1+\epsilon}{2}$. Observe that for the Tsirelson bound $\epsilon = \frac{1}{\sqrt{2}} \simeq 0.707$, $C_{\text{box}} = 0$. This lower bound is an implication of Information Causality.

following problem. Assume that Alice and Bob are observing i.i.d. copies of a and b respectively. Alice is interested in creating i.i.d. copies of x whereas Bob is interested in creating i.i.d. copies of y within a vanishing total variation distance from the distribution $p(x, y, a, b)$. To accomplish this, the two nodes exchange messages. The goal is to find the minimum communication rates, i.e. number of bits exchanged per i.i.d. observation of a, b . A formal definition of the simulation problem can be found in [23].

Information theoretic treatment of the above problem assumes that common randomness is provided to the parties at a given rate. However, in our setting we assume that the two parties are provided with infinite common randomness. It is shown in [24] that the minimum one-way communication rate from Alice to Bob with infinite preshared randomness when both b and x are constant random variables, is equal to $I(a; y)$. The exact formula for the communication rate has been obtained by Yassaee [25]. The answer is the maximum of $I(a; u|b)$ over all classical random variables u determined by $p(u|a, b, x, y)$ such that the joint distribution $p(u, a, b, x, y)$ factorizes as $p(u, a, b, x, y) = p(a, b)p(u|a)p(x|u, a)p(y|u, b)$. Its proof is beyond the scope of this paper [31].

For imperfect PR-boxes defined above, with uniform distribution on inputs ($p(a, b) = p(a)p(b) = \frac{1}{4}$), independence of a and b implies that $I(a; u|b) = I(a; u)$. Moreover, u can be taken to be a binary random variable using the Fenchel extension of the Caratheodory theorem. Then computing the optimal rate for every ϵ is a straightforward optimization problem. Fig. 3 gives the one-way communication cost of winning the CHSH game with probability p .

VI. CONCLUSION

We generalized and improved our understanding of Information Causality by connecting it to the Gray-Wyner problem. We showed that, assuming a new property on the underlying physical theory (Accessibility of Mutual Information), the classical Gray-Wyner region completely characterizes the game of Information Causality. That is, we provide an infinite number of inequalities one of which is Information Causality. Our assumption of Accessibility of Mutual Information is obvious in the classical case and holds in the quantum theory. Since the underlying assumptions of Information Causality have been recently studied for a general physical theory [26, 27], it is a natural question whether AMI can be verified for other physical theories. It is also interesting to see whether other Tsirelson's inequalities (besides the CHSH example) can be derived from these new inequalities. Moreover two-way communication protocols and multiparty scenarios are natural extensions of the Information Causality game and can be studied in a general physical theory.

We also studied the problem of simulating non-local correlations. We showed that Information Causality gives a bound on the rate of required communication assuming preshared entanglement. Moreover, we reported a formula to compute the optimal rate of required communication assuming infinite preshared randomness.

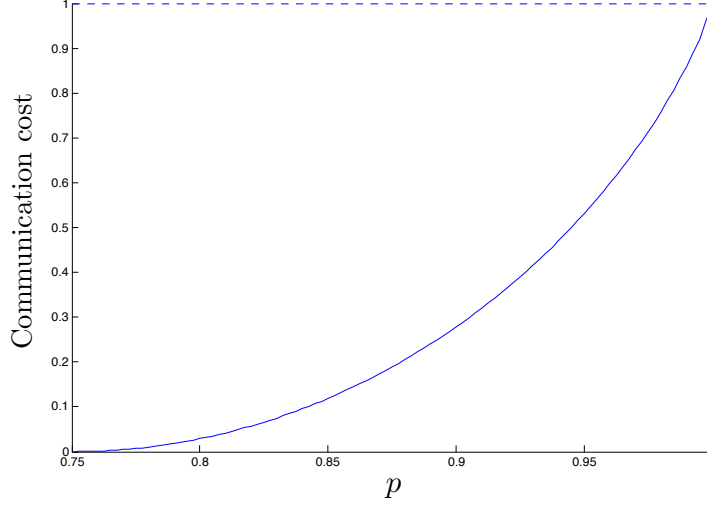


FIG. 3: The one-way communication cost of winning the CHSH game with probability $p = \frac{1+\epsilon}{2}$ assuming preshared randomness.

Acknowledgments

Authors are thankful to Robert Koenig for helping with some technical questions, and Peter W. Shor for suggesting the problem of simulating non-locality with preshared entanglement. This research was in part supported by a grant from IPM (No. CS1390-3-01), and the Iranian National Science Foundation No. 89003743.

Appendix A: Part II of the proof of Theorem IV.1

Following the proof of Theorem IV.1 we need to show that there exists some random variable e^* such that

$$\begin{aligned} \frac{1}{n} I(\vec{a}^n; e_n) &= I(\vec{a}; e^*), \\ \frac{1}{n} H(a_i^n | e_n) &\geq H(a_i | e^*), \quad 1 \leq i \leq N. \end{aligned}$$

This in fact says that the n -letter version of the Gray-Wyner capacity region is equal to its single-letter, which holds because it is a general property of the capacity regions in information theory. Nonetheless, to be self-contained, we provide a proof.

As before α^n denotes n independent copies of α . For $1 \leq i < j \leq n$ we let $\alpha^{i:j}$ to be the $j - i + 1$ copies of α starting with the i -th one. For simplicity we denote $\alpha^{i:i}$ by $\alpha^{(i)}$. Moreover, for notational convenience we introduce $\vec{a}^* = (a_1^*, a_2^*, \dots, a_N^*)$ having the same joint distribution as $\vec{a} = (a_1, a_2, \dots, a_N)$ and find $p(e^* | \vec{a}^*)$ such that

$$\begin{aligned} \frac{1}{n} I(\vec{a}^n; e_n) &= I(\vec{a}^*; e^*), \\ \frac{1}{n} H(a_i^n | e_n) &\geq H(a_i^* | e^*), \quad 1 \leq i \leq N. \end{aligned}$$

Let q be a random variable uniform on $\{1, 2, \dots, n\}$ independent of all previous random variables. Define

$$\begin{aligned} e^* &= (e_n, \vec{a}^{1:q-1}, q), \\ a_i^* &= a_i^{(q)}, \end{aligned}$$

where by our convention $a_i^{(q)} = a_i^{q:q}$ is the q -th i.i.d. copy of a_i . Observe that $(a_1^*, a_2^*, \dots, a_N^*)$ has the same joint

distribution as (a_1, a_2, \dots, a_N) . Then using the chain rule we obtain

$$\begin{aligned}
 I(\vec{a}^*; e^*) &= I(\vec{a}^{(q)}; e_n, \vec{a}^{1:q-1}, q) \\
 &= I(\vec{a}^{(q)}; q) + I(\vec{a}^{(q)}; \vec{a}^{1:q-1} | q) + I(\vec{a}^{(q)}; e_n | q, \vec{a}^{1:q-1}) \\
 &= 0 + 0 + \frac{1}{n} \sum_{q=1}^n I(\vec{a}^{(q)}; e_n | \vec{a}^{1:q-1}) \\
 &= \frac{1}{n} I(\vec{a}^n; e_n).
 \end{aligned}$$

Furthermore, by the data processing inequality we have

$$H(a_i^* | e^*) = H(a_i^{(q)} | e_n, \vec{a}^{1:q-1}, q) = \frac{1}{n} \sum_{q=1}^n H(a_i^{(q)} | e_n, \vec{a}^{1:q-1}) \leq \frac{1}{n} \sum_{q=1}^n H(a_i^{(q)} | e_n, a_i^{1:q-1}) = \frac{1}{n} H(a_i^n | e_n).$$

-
- [1] J. S. Bell, On the Einstein-Podolsky-Rosen Paradox, *Physics* **1**, 3, 195–200 (1964).
 - [2] A. Aspect, Bell's inequality test: more ideal than ever, *Nature* **398**, 189–190, (1999).
 - [3] B. S. Cirel'son, Quantum Generalizations of Bell's Inequality, *Letters in Mathematical Physics* **4**, 93–100 (1980).
 - [4] S. Popescu and D. Rohrlich, Nonlocality as an axiom, *Foundations of Physics* **24**, 379–385 (1994).
 - [5] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Proposed experiment to test local hidden-variable theories, *Physical Review Letters* **23**, 880–884 (1969).
 - [6] M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Żukowski, Information Causality as a Physical Principle, *Nature* **461**, 1101–1104 (October 2009).
 - [7] R.M. Gray and A.D. Wyner, Source coding for a simple network, *The Bell System Technical Journal*, vol. 53, no. 9, 1681–1721 (November 1974).
 - [8] B. F. Toner and D. Bacon, Communication Cost of Simulating Bell Correlations, *Phys. Rev. Lett.* **91**, 187904 (2003).
 - [9] G. Brassard, H. Buhrman, N. Linden, A. A. Méthot, A. Tapp, and F. Unger, Limit on Nonlocality in Any World in Which Communication Complexity Is Not Trivial, *Phys. Rev. Lett.* **96**, 250401 (2006).
 - [10] G. Brassard, R. Cleve, and A. Tapp, The cost of exactly simulating quantum entanglement with classical communication, *Physical Review Letters* **83** 1874–1877 (1999).
 - [11] C. Branciard and N. Gisin, Quantifying the nonlocality of GHZ quantum correlations by a bounded communication simulation protocol, *Physical Review Letters* **107**, 020401 (2011).
 - [12] J. Degorre, S. Laplante, and J. Roland, Simulating quantum correlations as a distributed sampling problem, *Physical Review A* **72**, 062314 (2005).
 - [13] J. Degorre, M. Kaplan, S. Laplante, and J. Roland, The communication complexity of non-signaling distributions, *Quantum Information & Computation*, **11**(7 & 8): 649–676 (2011).
 - [14] Y. Shi and Y. Zhu, Tensor Norms and the Classical Communication Complexity of Nonlocal Quantum Measurement, *SIAM Journal on Computing* **38**, 753–766 (2008).
 - [15] O. Regev and B. Toner, Simulating Quantum Correlations with Finite Communication, *Simulating Quantum Correlations with Finite Communication*. *SIAM Journal on Computing*, **39** (4), 1562–1580 (2009).
 - [16] S.i W. Al-Safi and A. J. Short, Information causality, entropy and the inner product game, arXiv:1107.4031 (2011).
 - [17] O. C. O. Dahlsten, D. Lercher, and R. Renner, Tsirelson's bound from a Generalised Data Processing Inequality, arXiv:1108.4549 (2011).
 - [18] W. Liu, G. Xu, and B. Chen, The common information of N dependent random variables, *Proc. 48th Annual Allerton Conference on Communications, Control and Computing*, Monticello, IL (September 2010).
 - [19] A. El Gamal and Y.-H. Kim, Network Information Theory, Lecture notes, available online at <http://arxiv.org/abs/1001.3404>.
 - [20] G. Wang, Functional boxes, communication complexity and information causality, arXiv:1109.4988 (2011).
 - [21] V. Anantharam and V. Borkar, Common randomness and distributed control: A counterexample, *Systems and Control Letters*, **56**: 568–572 (2007).
 - [22] P. Cuff, T. Cover and H. Permuter, Coordination Capacity, *IEEE Transactions on Information Theory*, **56** (9): 4181–4206 (2010).
 - [23] A. Gohari, V. Anantharam, Generating Dependent Random Variables Over Networks, *Proceeding of the IEEE Information Theory Workshop*, Praty (2011).
 - [24] P. Cuff, Communication Requirements for Generating Correlated Random Variables, *Proc IEEE Int Symp Info Theory*, pp. 1393–1397, 2008.
 - [25] M. Yassaee, A. Gohari, M. Aref, in preparation, (2011).
 - [26] A. J. Short and S. Wehner, Entropy in general physical theories, *New Journal of Physics* **12**, 033023 (2010).

- [27] H. Barnum, J. Barrett, L. O. Clark, M. Leifer, R. Spekkens, N. Stepanik, A. Wilce, and R. Wilke, Entropy and Information Causality in General Probabilistic Theories, *New Journal of Physics* **12**, 033024 (2010).
- [28] Wyner's common information is an operational way of defining the common part of two dependent random variables. It is given by $J(a_1, a_2) = \min_{a_1 \rightarrow u \rightarrow a_2} I(u; a_1 a_2)$ and satisfies $I(a_1; a_2) \leq J(a_1, a_2) \leq H(a_1, a_2)$.
- [29] \mathcal{R} is closed because the range of the auxiliary random variable e in the definition of \mathcal{R} is finite and the entropy function is continuous.
- [30] It is easy to see that the data processing inequality holds in the conditional form as well.
- [31] Briefly speaking the converse follows from the traditional Gallager type auxiliary identification with the choice of auxiliary being identical to one given in [23]. The achievability follows from a coloring lemma on the typical sequences of u, a .